DOCUMENT RESUME

ED 410 296 TM 027 169

AUTHOR Fitzgerald, T. J.; And Others
TITLE Medical Student Self-Assessment Abilities: Accuracy and Calibration.
PUB DATE Mar 97
NOTE 7p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Estimation (Mathematics); Higher Education; Knowledge Level; Medical Education; *Medical Students; Performance Factors; Scores; *Self Evaluation (Individuals); *Test Results
IDENTIFIERS *Accuracy; *Calibration

ABSTRACT

Practicing physicians must assess their own knowledge and skills continually to ensure that they are current with new medical procedures and advances. The ability to assess one's own performance was studied for medical students, who were asked to estimate their performance on course quizzes. An item that asked students to estimate their correct percentage on the examination was added to five weekly quizzes in the winter term. The number of students providing self-evaluations ranged from 137 to 164 in a class of 170 students. Overall, students were quite accurate in predicting their test scores. However, their assessments were not well calibrated in that the variations in their estimates did not correlate with variations in their actual performance. The high level of accuracy in their self-assessments does indicate well-developed self-assessment skills, but whether the ability to evaluate one's own performance carries over into areas less familiar than examinations remains to be determined. (Contains one figure and five references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM027169

ED 410 296

Medical Student Self-assessment Abilities: Accuracy and Calibration

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
James Fitzgerald

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

T.J. Fitzgerald
L.D. Gruppen
C. White
W.K. Davis

TM027169

This paper is prepared for the Annual Meeting of the American Educational Research Association in Chicago, IL

BEST COPY AVAILABLE

Medical Student Self-assessment Abilities: Accuracy and Calibration

J.T. Fitzgerald, L. D. Gruppen, C. White, W. K. Davis
University of Michigan Medical School

J.T. Fitzgerald
G1111 Towsley
Ann Arbor, MI
48109-0201

Practicing physicians must continually assess their own knowledge and skills to ensure currency with new medical procedures and advances. Ideally, they address any recognized gaps in their knowledge base by identifying and using relevant educational resources. However, this image of the self-directed physician-learner has received much more attention as a proposed goal and ideal of medical education than it has in the theoretical or empirical realm.

The breadth of the self-directed learning concept and the complex interaction of constituent abilities and individual characteristics makes it a challenging area to study. However, one component of self-directed learning that is relatively amenable to investigation is a physician's ability to assess accurately her professional performance.

Prior research into self-assessment, though limited, has yielded several discoveries. There appears to be at least a modest developmental component in medical students' ability to evaluate themselves (and peers), which lags behind their ability to perform specific skills.[1] Self-assessment may also be modifiable through education as suggested by findings that students' self-assessment skills increase slightly over the course of education,[2] and that students' evaluation criteria become more stringent with experience.[3] However, even if self-assessment is a learnable skill, there is a strong probability that much of this learning has taken place in childhood and that it may be rather fixed by the time students enter medical school.[4] The limited evidence of improvement in self-assessment skills during medical training may reflect the relatively fixed character of adult self-assessment, but it may also reflect the fact that students and residents receive relatively little practice in self-assessment.[5]

Existing research on self-assessment has examined relatively broad, general skills such as the ability to accurately define areas of strength or weakness in one's knowledge or skills.[5] Complementing this broad approach is a more focused approach that concentrates on a task or case level, e.g., "How well did I do in Mrs. Peterson's gall bladder surgery?" Task-focused, as opposed to general skills-focused self-assessment, can probably be more readily studied in medical students who have ample opportunity to self assess performance on specific tasks (i.e., tests and quizzes) and less opportunity to assess more general clinical skills and knowledge. This study examines self-assessment abilities among first-year medical students on a task-specific basis; specifically, how well they estimate their performance on course quizzes.

Methods

To determine first-year University of Michigan medical students' self-assessment skills, a single item was added to five weekly quizzes in the winter term. The added item asked the students to "Please estimate your percent correct on this exam (0% - 100%)." Student responses to this item provided data on the same percent-correct scale as their actual performance score. The quizzes consisted of 35- to 40-items from all first-year courses taught early in the winter term (Embryology, Histology, Molecular and Cellular Biology, Physiology). Each quiz took approximately 60 minutes to complete. The number of students providing self estimates ranged from 137 to 164 in a class of 170 students.

Accuracy and Calibration

In addition to the conceptual complexities of studying self-assessment, there are some methodological issues that merit attention. Most prior studies of self-assessment have correlated student self ratings with similar ratings of the student by faculty. Conceptually, there is a question about the basis on which such evaluations are made by the student and the faculty member. Typically, these ratings are implicitly or explicitly designed to evaluate the student in the context of the class. Although faculty may have the experiential basis for comparing a student to her peers, students cannot be expected to possess similar knowledge about the performance,

TM027169

skills, or knowledge of the other members of the class. Thus, students' estimates based on the entire group may have limited validity.

Even if a student's rank in comparison to peers could be validly obtained from students, it would probably not be very useful in directing self-education, which should (arguably) focus more on individual strengths and weakness than group standards. Thus, rather than measuring self-assessment through a 'between subjects' procedure, a 'within subjects' model may be more valid and informative. Such a model would call for students to evaluate relative strengths and weaknesses within themselves, without comparison to others. This avoids the problem of different interpretations of evaluation criteria and enables the students to use their own experience and past performance as a comparative standard.

The 'within subjects' model guided the self-assessment task used in this study, which asked students to estimate how well they performed on a quiz they had just completed. This procedure essentially replaces the faculty's evaluation with an 'objective' evaluation and may shift the basis for explaining good or poor self-assessment toward metacognitive skills and psychological factors (e.g., locus of control, causal attributions) rather than toward the nature of professional or expert evaluation.

Two 'within-subjects' measures are defined to examine student self-assessment. The first is the difference between each student's estimated performance and actual score on a given quiz. This *accuracy* variable provides information about whether a student has over- or underestimated their performance, and by how much. When averaged over multiple observations, the mean of these differences for a given student provides a single value that reflects the amount and direction of any overall 'bias' in how this student self-assesses her performance.

The second variable assesses the correlation between a student's estimated and actual performances over multiple observations. This correlation summarizes how well *calibrated* a student's self-assessment is with her actual performance, i.e., the extent to which variations in a student's estimates parallel variations in actual performance. A correlation of +1.00 indicates a student who provides her highest estimated performance on the quiz on which she does the best (of the other quizzes in the analysis) and her lowest estimate on the quiz on which she does the worst. A perfect negative correlation indicates a student who thought he did the best on the quiz on which he actually performed the worst, and vice versa. Note that the correlation is not influenced by differences between the values of the estimated and actual scores (i.e., accuracy) but only reflects covariation.

In order to evaluate the impact of academic ability on these self-assessment variables, accuracy and calibration were covaried with both MCAT score and mean score over all five quizzes.

Results

Self-assessment Accuracy

The average of student accuracy scores (predicted minus actual) was -1.05 ± 5.00, indicating that overall, students were quite accurate in estimating their test scores.

The correlation between student accuracy and MCAT score was only 0.01. However, a significant negative correlation was found between student accuracy and their average quiz score ($r=-.42$, $p<.001$), indicating that higher performance on the quizzes was associated with less accurate estimates of performance. An analysis of mean accuracy scores by level of quiz performance clarified this relationship. The students at the bottom 25% of the distribution of average quiz scores tended to overestimate their performance (mean accuracy=1.73) while the other three quartiles underestimated their performance. Students with the top 25% scores underestimated their performance by the largest amount (mean accuracy=-3.35); the 51% to 75% group averaged -2.50, while the 26% to 50% averaged -0.10.

Self-assessment Calibration

Wide variation was found in student calibration values, ranging from very high (1.00) to very low (-.98, see Figure 1). Furthermore, the average individual calibration was much lower (r=.35)

than the group-based correlation found between the average quiz scores and average self-assessment (r=.65).

The correlation between student calibration values and MCAT scores was low and negative(r=-.22, p=.02), indicating that higher student calibration was weakly associated with lower MCAT scores. The correlation between student calibration and average quiz score was not significant (r=-.12, p=.14).

The two measures of student self-assessment, accuracy and calibration, proved to be independent (r=-.003, p=.97).

Discussion

These findings suggest that, as a group, first-year medical students are quite accurate in assessing their performance on course quizzes. Their self-assessments are not, however, well-calibrated; the variations in their estimates do not correlate with the variations in their actual performance.

The high level of accuracy in these students' self-assessments (within 1% of their actual performance) is striking, and suggests well-developed self-assessment skills. However, there are several caveats to consider. Estimating performance on knowledge tests, such as the quizzes in this study, is something with which the students have had considerable experience. Whether similar levels of self-assessment accuracy will be observed in more novel tasks, such as clinical performance evaluations and standardized patient interactions, remains to be determined. Additionally, these were not the first quizzes of the year, so students had an opportunity to 'tune' their self-assessment by the time we gathered the estimates presented here.

The observed relationship between self-assessment accuracy and quiz performance likely reflects a floor and ceiling artifact in that any inaccuracies in self-assessment of students who perform at the top of the distribution are most likely to be underestimates (ceiling effect) and the inaccuracies of students at the lower quartile of the distribution are more likely to be overestimates (floor effects).

The wide range of values for student calibration is intriguing and suggests the need for further investigation of the implications of this variable, as well as other performance and psychological variables that may be associated with it. The fact that self-assessment calibration and accuracy are not correlated with each other suggests that these are two independent characteristics of self-assessment which may have distinct contributions to make to the understanding of this process.

Methodologically, it is worth noting the contrasting results obtained from using an individual-based vs. a group-based correlation between actual and estimated performance. When the students' actual and predicted scores were averaged within students to produce a single pair of values for each student and then correlated over the group, which is the kind of correlation produced in prior studies, the result is a moderately high correlation of .65. However, when the correlations are done on an individual basis over the five quizzes, and then these individual correlations averaged, the result is a mean correlation of .35. Not only do these procedures show different aspects of the data, we would argue that the individual-based correlations have considerable value in identifying individuals who are well- or poorly-calibrated. This is impossible to obtain from a group-based correlation.

The ability to assess skills and knowledge levels correctly is essential for the practicing physician. It is important not only in seeking continuing medical education, but also in referral behavior (when to treat and when to send to a specialist). Assessing and improving this skill in medical students is essential to maintaining the lifelong competency of physicians and high-quality care for their patients.

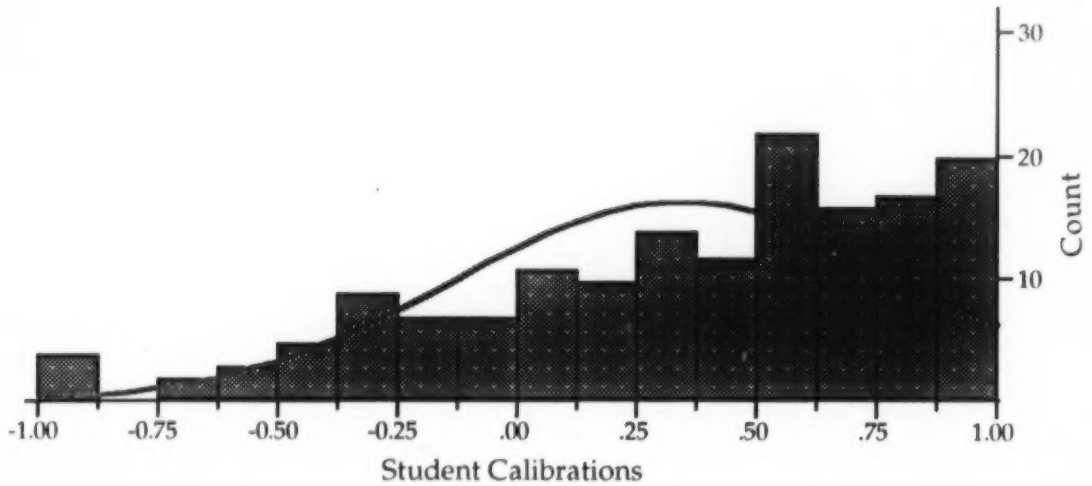
Acknowledgment

We would like to acknowledge Glenn Regehr, Ph.D., University of Toronto, for his assistance in helping us to better understand the methodological issues of self-assessment.

References

1. Calhoun JG, Woolliscroft JO, Hockman EM, Wolf FM, Davis WK. Evaluating medical student clinical skill performance: Relationships among self, peer, and expert ratings. *Proceedings of the 23rd annual conference on Research in Medical Education* 1984;23:205-210.
2. Arnold L, Willoughby TL, Calkins EV. Self-evaluation in undergraduate medical education: A longitudinal perspective. *Journal of Medical Education* 1985;60:21-28.
3. Calhoun JG, Ten Haken JD, Woolliscroft JO. Medical students' development of self- and peer-assessment skills: A longitudinal study. *Teaching and Learning in Medicine* 1990;2:25-29.
4. Woolliscroft JO, Ten Haken J, Smith J, Calhoun JG. Medical students' clinical self-assessments: Comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Academic Medicine* 1993:68;285-294.
5. Sclabassi SE, Woelfel SK. Development of self-assessment skills in medical students. *Medical Education* 1984;84:226-231.

Figure 1
Distribution of individual self-assessment calibration values
n=159



Abstract

This study examined the ability of first-year medical students to self-assess their performance on five class quizzes. Two variables were defined to measure student self-assessment, accuracy (the difference between the students' estimated and actual score) and calibration (the correlation between a student's estimated and actual scores). The results indicated that the students were fairly accurate in assessing their performance on the quizzes, but that their assessments were not calibrated. Furthermore, the measures of calibration and accuracy did not correlate suggesting each is measuring different aspects of self-assessment.